

CLAIMS

1. An apparatus for determining, based on speech waveform data, a portion reliably representing a feature of the speech waveform, comprising:
 - 5 extracting means for calculating, from said data, distribution of an energy of a prescribed frequency range of said speech waveform on a time axis, and for extracting, among various syllables of said speech waveform, a range that is generated stably by a source of said speech waveform, based on the distribution and pitch of said speech waveform;
 - 10 estimating means for calculating, from said data, distribution of spectrum of said speech waveform on the time axis, and for estimating, based on the spectral distribution on the time axis, a range of said speech waveform of which change is well controlled by said source; and
 - 15 means for determining that range which is extracted by said extracting means as the range generated stably by said source and of which speech waveform is estimated by said estimating means to be well controlled by said source, as a highly reliable portion of said speech waveform.
- 20 2. The apparatus according to claim 1, wherein
said extracting means includes
voiced/unvoiced determining means for determining, based on said data, whether each segment of said speech waveform is a voiced segment or not,
- 25 means for separating said speech waveform into syllables at a local minimum of said waveform of energy distribution of the prescribed frequency range of said speech waveform on the time axis; and
- 30 means for extracting that range of said speech waveform which includes, in each syllable, an energy peak in that syllable within the segment determined to be a voiced segment by said voiced/unvoiced determining means and in which the energy of the prescribed frequency range is not lower than a prescribed threshold value.

3. The apparatus according to claim 1, wherein
said estimating means includes
linear predicting means for performing linear prediction analysis on
said speech waveform and outputting an estimated value of formant
frequency;

5 first calculating means for calculating, using said data, distribution
of non-reliability of the estimated value of formant frequency provided by
said linear predicting means on the time axis;

10 second calculating means for calculating, based on an output from
said linear predicting means, distribution on the time axis of local variance
of spectral change on the time axis of said speech waveform; and

15 means for estimating, based both on said distribution on the time
axis of non-reliability of the estimated value of formant frequency
calculated by said first calculating means and on said distribution on the
time axis of local variance of spectral change in said speech waveform
calculated by said second calculating means, a range in which change in
the speech waveform is well controlled by said source.

20 4. The apparatus according to claim 1, wherein
said determining means includes

means for determining, as a highly reliable portion of said speech
waveform, a range included in the range extracted by said extracting
means, within the range of which change in speech waveform is estimated
by said estimating means to be well controlled by said source.

25 5. A quasi-syllabic nuclei extracting apparatus for separating a
speech signal into quasi-syllables and extracting a nuclear portion of each
quasi-syllable, comprising:

30 voiced/unvoiced determining means for determining whether each
segment of the speech signal is voiced or not;

means for separating said speech signal into quasi-syllables at a
local minimum of time-distribution waveform of an energy of a prescribed
frequency range of said speech signal; and

means for extracting that range of said speech signal which includes energy peak in each quasi-syllable, determined by said voiced/unvoiced determining means to be a voiced segment and of which energy of the prescribed frequency range is not lower than a prescribed threshold value,
5 as the nuclei of quasi-syllable.

6. The quasi-syllabic nuclei extracting apparatus according to claim
5, wherein

said extracting means includes

10 means for extracting that range of said speech signal which includes an energy peak in each pseudo-syllable within the segment determined to be a voiced segment by said voiced/unvoiced determining means and in which the energy of said prescribed frequency range is not lower than a prescribed threshold value as the nuclei of quasi-syllable.

15 7. An apparatus for determining a portion representing, with high reliability, a feature of a speech signal, comprising:

linear predicting means for performing linear prediction analysis on said speech signal;

20 first calculating means for calculating, based on an estimated value of formant provided by said linear predicting means and on said speech signal, distribution on time axis of non-reliability of the formant estimated value;

25 second calculating means for calculating, based on the result of linear prediction analysis by said linear predicting means, distribution on time axis of local variance of spectral change in said speech signal; and

30 means for estimating, based on the distribution on time axis of the non-reliability of the estimated value of formant frequency calculated by said first calculating means, and on the distribution on time axis of local variance of spectral change in said speech waveform calculated by said second calculating means, a range in which the change in said speech waveform is well controlled by said source.

8. A program product causing, when executed on a computer, said computer to operate as an apparatus for determining, based on speech waveform data, a portion reliably representing a feature of the speech waveform, said apparatus comprising:

5 extracting means for calculating, from said data, distribution of an energy of a prescribed frequency range of said speech waveform on a time axis, and for extracting, among various syllables of said speech waveform, a range that is generated stably by a source of said speech waveform, based on the distribution and pitch of said speech waveform;

10 estimating means for calculating, from said data, distribution of spectrum of said speech waveform on the time axis, and for estimating, based on the spectral distribution on the time axis, a range of said speech waveform of which change is well controlled by said source; and

15 means for determining that range which is extracted by said extracting means as the range generated stably by said source and of which speech waveform is estimated by said estimating means to be well controlled by said source, as a highly reliable portion of said speech waveform.

20 9. The program product according to claim 8, wherein
said extracting means includes

voiced/unvoiced determining means for determining, based on said data, whether each segment of said speech waveform is a voiced segment or not,

25 means for separating said speech waveform into syllables at a local minimum of said waveform of energy distribution of the prescribed frequency range of said speech waveform on the time axis; and

30 means for extracting that range of said speech waveform which includes, in each syllable, an energy peak in that syllable within the segment determined to be a voiced segment by said voiced/unvoiced determining means and in which the energy of the prescribed frequency range is not lower than a prescribed threshold value.

10. The program product according to claim 8, wherein
said estimating means includes
linear predicting means for performing linear prediction analysis on
said speech waveform and outputting an estimated value of formant
frequency;

5 first calculating means for calculating, using said data, distribution
of non-reliability of the estimated value of formant frequency provided by
said linear predicting means on the time axis;

10 second calculating means for calculating, based on an output from
said linear predicting means, distribution on the time axis of local variance
of spectral change on the time axis of said speech waveform; and

15 means for estimating, based both on said distribution on the time
axis of non-reliability of the estimated value of formant frequency
calculated by said first calculating means and on said distribution on the
time axis of local variance of spectral change in said speech waveform
calculated by said second calculating means, a range in which change in
the speech waveform is well controlled by the source.

20 11. The program product according to claim 8, wherein
said determining means includes

means for determining, as a highly reliable portion of said speech
waveform, a range included in the range extracted by said extracting
means, within the range of which change in speech waveform is estimated
by said estimating means to be well controlled by said source.

25 12. A program product causing, when executed on a computer, said
computer to operate as a quasi-syllabic nuclei extracting apparatus for
separating a speech signal into quasi-syllables and extracting a nuclear
portion of each quasi-syllable, said quasi-syllabic nuclei extracting
apparatus comprising:

30 voiced/unvoiced determining means for determining whether each
segment of the speech signal is voiced or not;
means for separating said speech signal into quasi-syllables at a

local minimum of time-distribution waveform of an energy of a prescribed frequency range of said speech signal; and

means for extracting that range of said speech signal which includes energy peak in each quasi-syllable, determined by said voiced/unvoiced determining means to be a voiced segment and of which energy of the prescribed frequency range is not lower than a prescribed threshold value, as the nuclei of quasi-syllable.

13. A program product causing a computer to operate as an apparatus for determining a portion representing, with high reliability, a feature of a speech signal, said apparatus comprising:

linear predicting means for performing linear prediction analysis on said speech signal;

15 first calculating means for calculating, based on an estimated value of formant provided by said linear predicting means and on said speech signal, distribution on time axis of non-reliability of the formant estimated value;

20 second calculating means for calculating, based on the result of linear prediction analysis by said linear predicting means, distribution on time axis of local variance of spectral change in said speech signal; and

25 means for estimating, based on the distribution on time axis of the non-reliability of the estimated value of formant frequency calculated by said first calculating means, and on the distribution on time axis of local variance of spectral change in said speech waveform calculated by said second calculating means, a range in which the change in said speech waveform is well controlled by said source.

14. A method of determining, based on speech waveform data, a portion reliably representing a feature of the speech waveform, comprising the steps of:

30 calculating, from said data, distribution of an energy of a prescribed frequency range of said speech waveform on a time axis, and extracting, among various syllables of said speech waveform, a range that is generated

stably by a source of said speech waveform, based on the distribution and pitch of said speech waveform;

5 calculating, from said data, distribution of spectrum of said speech waveform on the time axis, and estimating, based on the spectral distribution on the time axis, a range of said speech waveform of which change is well controlled by said source; and

10 determining that range which is extracted in said extracting step as the range generated stably by said source and of which speech waveform is estimated in said estimating step to be well controlled by said source, as a highly reliable portion of said speech waveform.

15 15. The method according to claim 14, wherein
said extracting step includes the steps of
determining, based on said data, whether each segment of said
speech waveform is a voiced segment or not,

detecting a local minimum of said waveform of energy distribution of the prescribed frequency range of said speech waveform on the time axis, and separating said speech waveform into syllables at the local minimum; and

20 extracting that range of said speech waveform which includes, in each syllable, an energy peak in that syllable within the segment determined to be a voiced segment by said voiced/unvoiced determining means and in which the energy of the prescribed frequency range is not lower than a prescribed threshold value.

25 16. The method according to claim 14, wherein
said estimating step includes
performing linear prediction analysis on said speech waveform and outputting an estimated value of formant frequency;

30 calculating, using said data, distribution of non-reliability of the estimated value of formant frequency on the time axis provided in said step of outputting the estimated value;

calculating, based on the calculated distribution of non-reliability of

the estimated value of formant frequency on the time axis, distribution on the time axis of local variance of spectral change on the time axis of said speech waveform; and

estimating, based both on said calculated distribution on the time axis of non-reliability of the estimated value of formant frequency and on said calculated distribution on the time axis of local variance of spectral change in said speech waveform, a range in which change in the speech waveform is well controlled by said source.

10 17. The method according to claim 14, wherein
said determining step includes the step of
determining, as a highly reliable portion of said speech waveform, a
range included in the range extracted in said extracting step, within the
range of which change in speech waveform is estimated in said estimating
15 step to be well controlled by said source.

18. A method of separating a speech signal into quasi-syllables and
extracting a nuclear portion of each quasi-syllable, comprising the steps of:
determining whether each segment of the speech signal is voiced or
not;

20 separating said speech signal into quasi-syllables at a local minimum
of time-distribution waveform of an energy of a prescribed frequency range
of said speech signal; and
extracting that range of said speech signal which includes energy
25 peak in each quasi-syllable, determined in said voiced/unvoiced
determining step to be a voiced segment and of which energy of the
prescribed frequency range is not lower than a prescribed threshold value,
as the nuclei of quasi-syllable.

30 19. The method according to claim 18, wherein
said extracting step includes the step of
extracting that range of said speech signal which includes an energy
peak in each pseudo-syllable within the segment determined to be a voiced

segment in said voiced/unvoiced determining step and in which the energy of said prescribed frequency range is not lower than a prescribed threshold value as the nuclei of quasi-syllable.